US 20080159425A1

(19) **United States**

(12) **Patent Application Publication**

(10) **Pub. No.: US 2008/0159425 A1**

(43) **Pub. Date: Jul. 3, 2008**

Khojastepour et al.

(54) **DESIGN OF MULTI-USER DOWNLINK LINEAR MIMO PRECODING SYSTEMS**

(76) Inventors: **Mohammad A. Khojastepour**, North Brunswick, NJ (US); **Xiaodong Wang**, New York, NY (US); **Mohammad Madihian**, Plainsboro, NJ (US)

Correspondence Address:
NEC LABORATORIES AMERICA, INC.
4 INDEPENDENCE WAY, Suite 200
PRINCETON, NJ 08540

(21) Appl. No.: **12/002,874**

(22) Filed: **Dec. 19, 2007**

Related U.S. Application Data

(60) Provisional application No. 60/870,930, filed on Dec. 20, 2006.

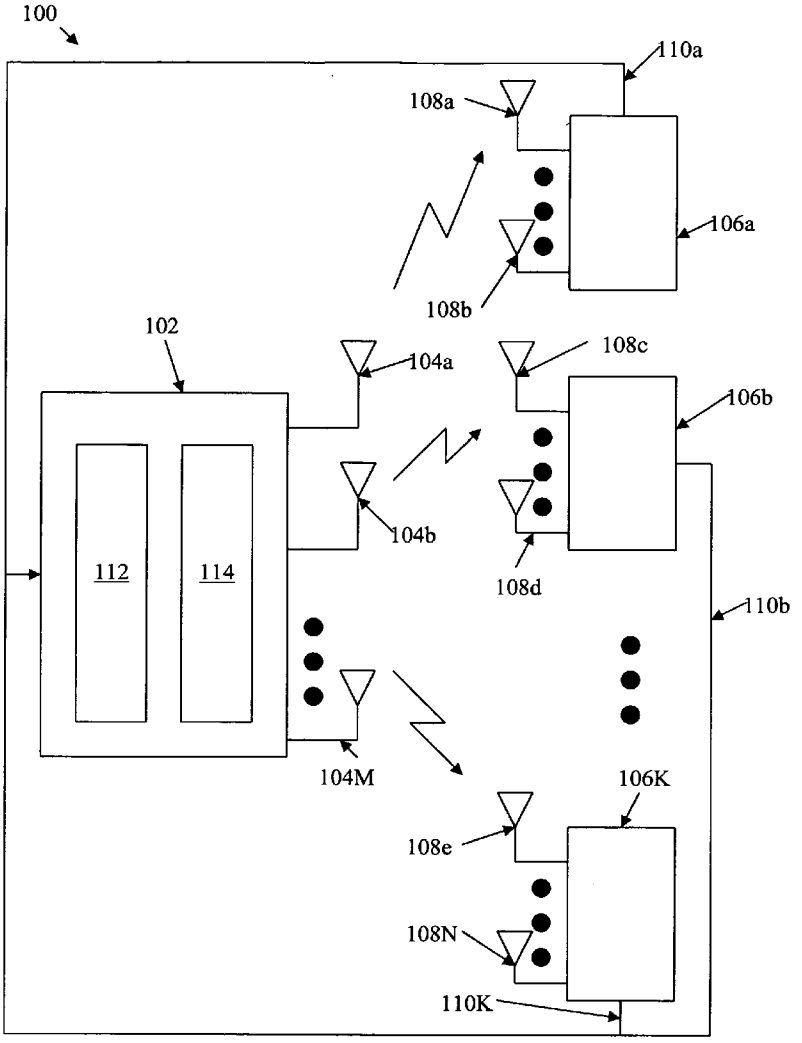
Publication Classification

(51) **Int. Cl.**
H04L 27/28 (2006.01)

(52) **U.S. Cl.** **375/260**

(57) **ABSTRACT**

Multi-user (MU-) MIMO systems with quantized feedback are designed to maximize the sum-rate via scheduling and linear precoding. To maximize throughput over the network, quantized CSIT is sent through a low-rate feedback link feedback from a plurality of users back to a base station. The base station then determines a subset of the plurality of users to transmit one or more signals to based on the received feedback and determines a preceding matrix based on the received feedback from the plurality of users wherein the precoding matrix maximizes a sum-rate throughput for the subset of the plurality of users. Additionally, based on the received feedback, the base station designs a quantization codebook. This codebook may be designed off-line and/or online. The codebook and/or precoding matrix are used to transmit signals to the users.



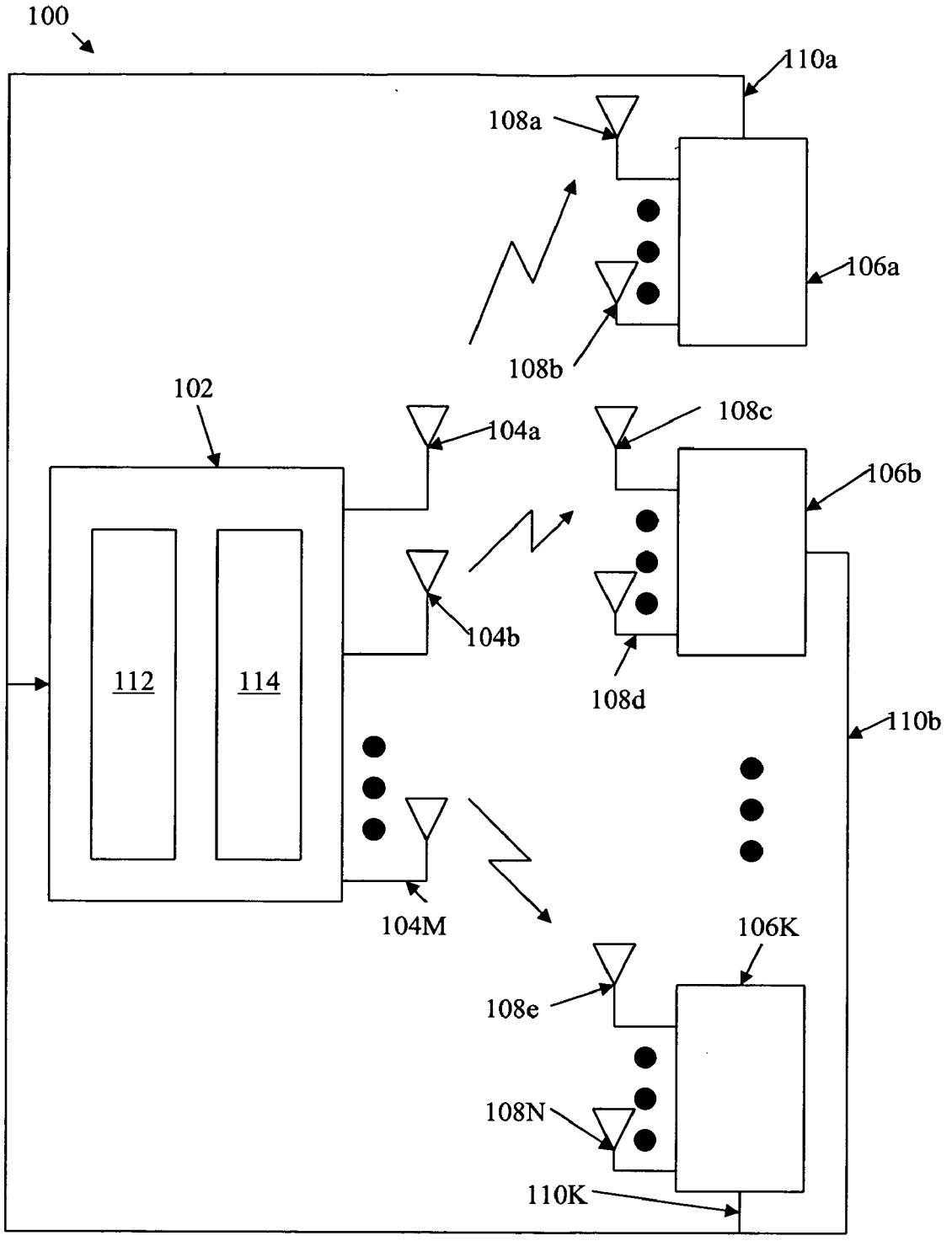

100
110a
108a
106a
108b
102
104a
108c
106b
104b
112
114
108d
110b
104M
106K
108e
108N
110K

FIG. 1

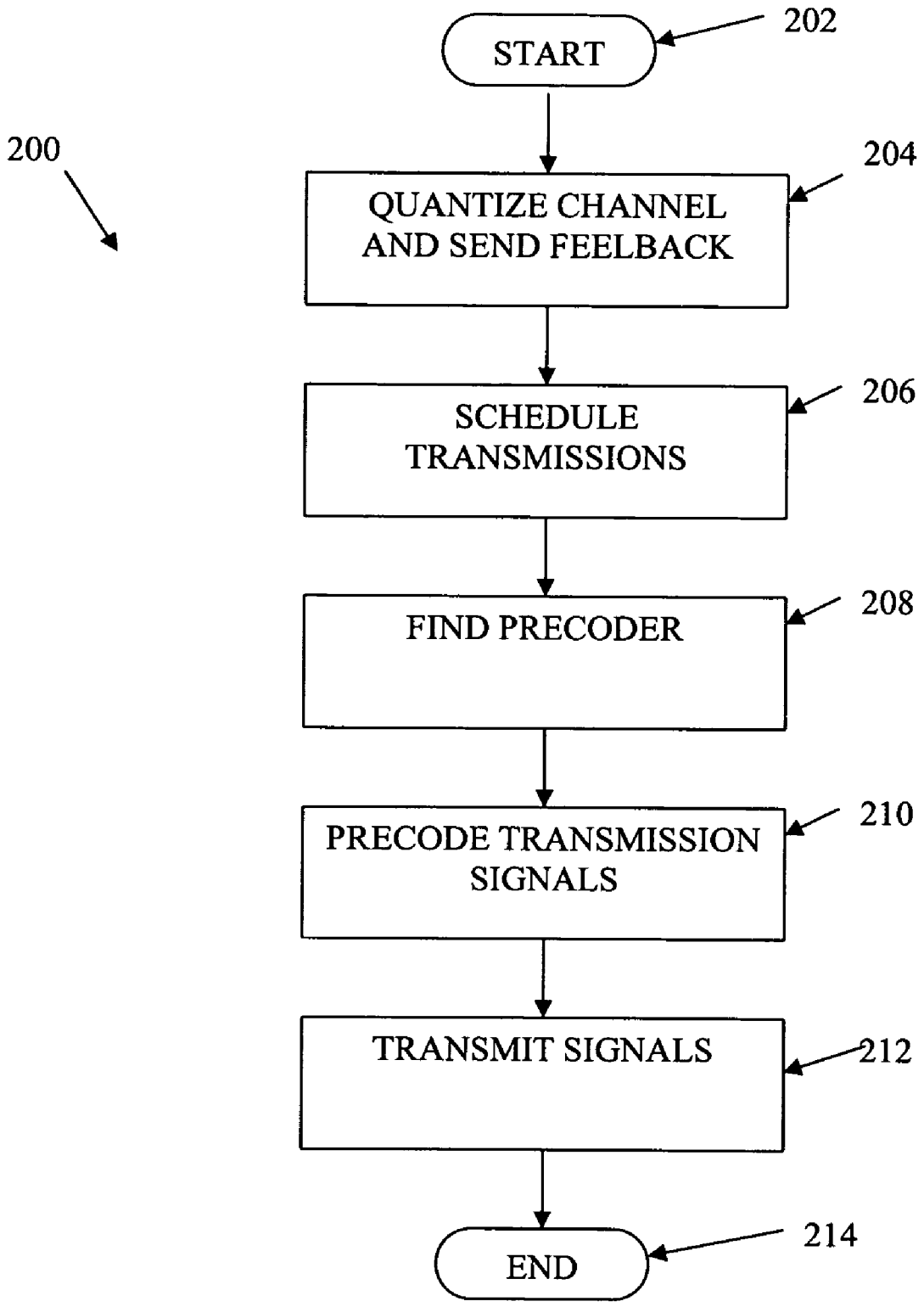

200
START
202
QUANTIZE CHANNEL AND SEND FEELBACK
204
SCHEDULE TRANSMISSIONS
206
FIND PRECODER
208
PRECODE TRANSMISSION SIGNALS
210
TRANSMIT SIGNALS
212
END
214

FIG. 2

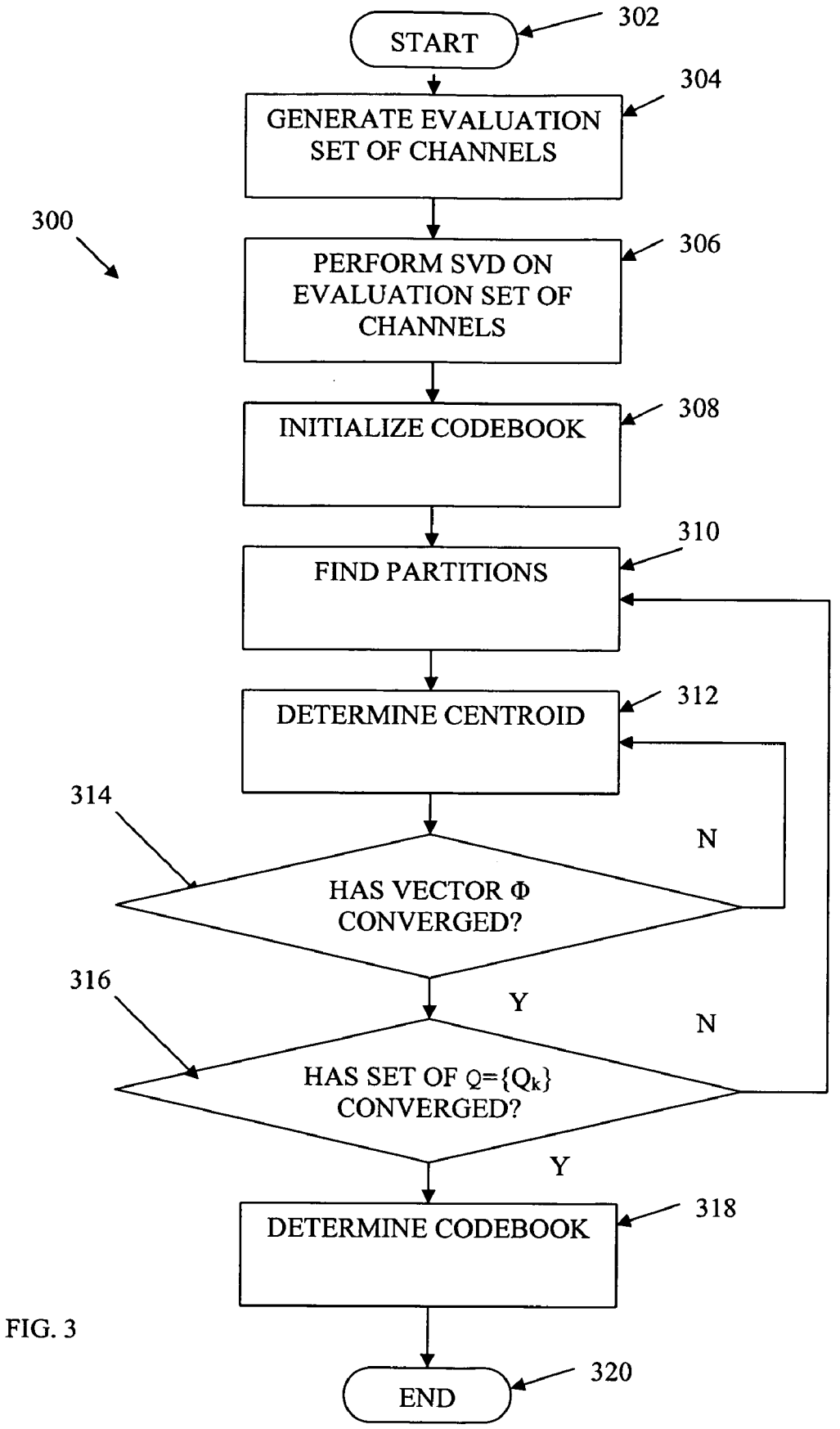

300
START
302
GENERATE EVALUATION SET OF CHANNELS
304
PERFORM SVD ON EVALUATION SET OF CHANNELS
306
INITIALIZE CODEBOOK
308
FIND PARTITIONS
310
DETERMINE CENTROID
312
HAS VECTOR Φ CONVERGED?
314
N
Y
HAS SET OF Q={Qk} CONVERGED?
316
N
Y
DETERMINE CODEBOOK
318
END
320

FIG. 3

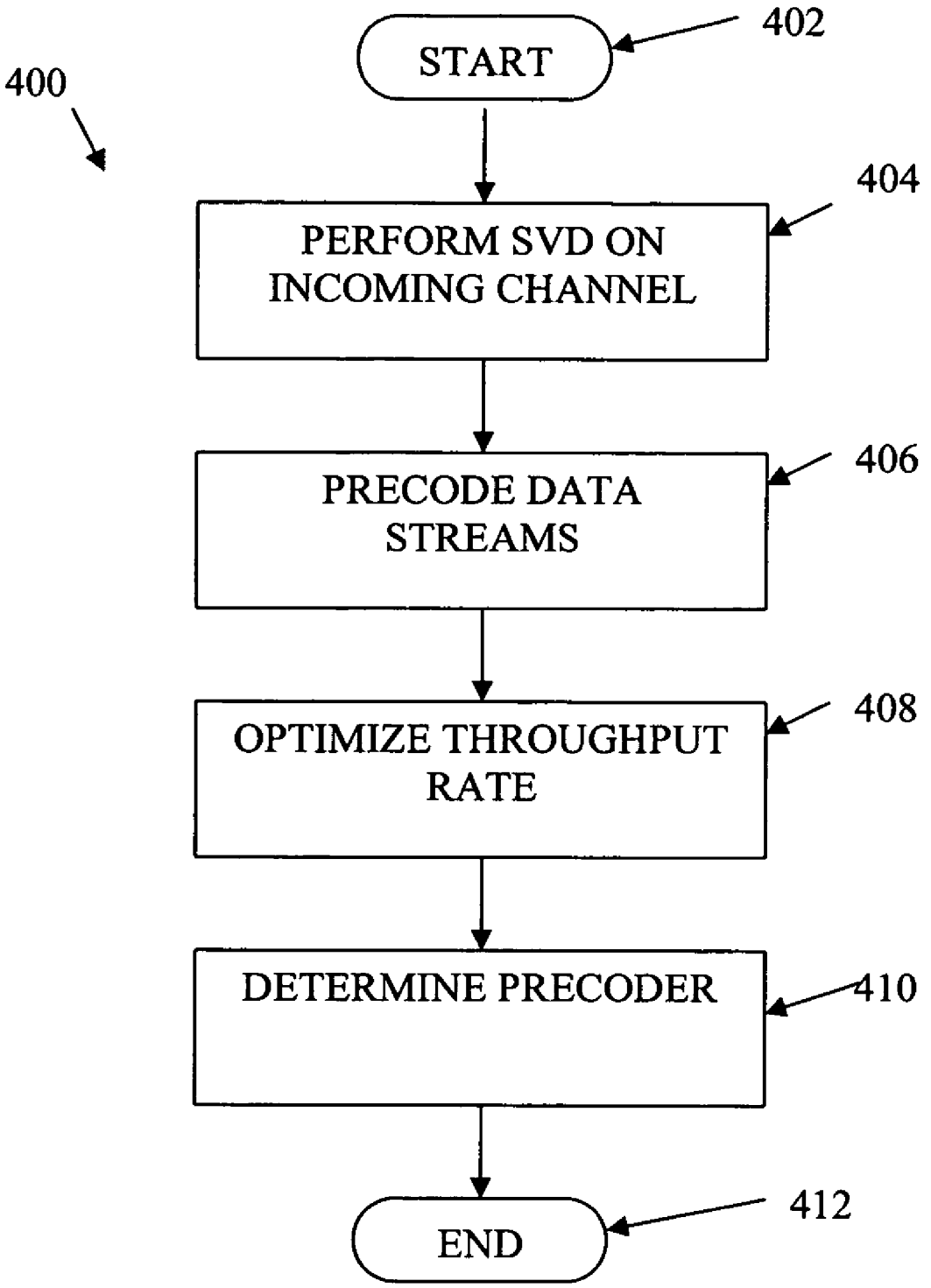

400
START
402
PERFORM SVD ON INCOMING CHANNEL
404
PRECODE DATA STREAMS
406
OPTIMIZE THROUGHPUT RATE
408
DETERMINE PRECODER
410
END
412

FIG. 4

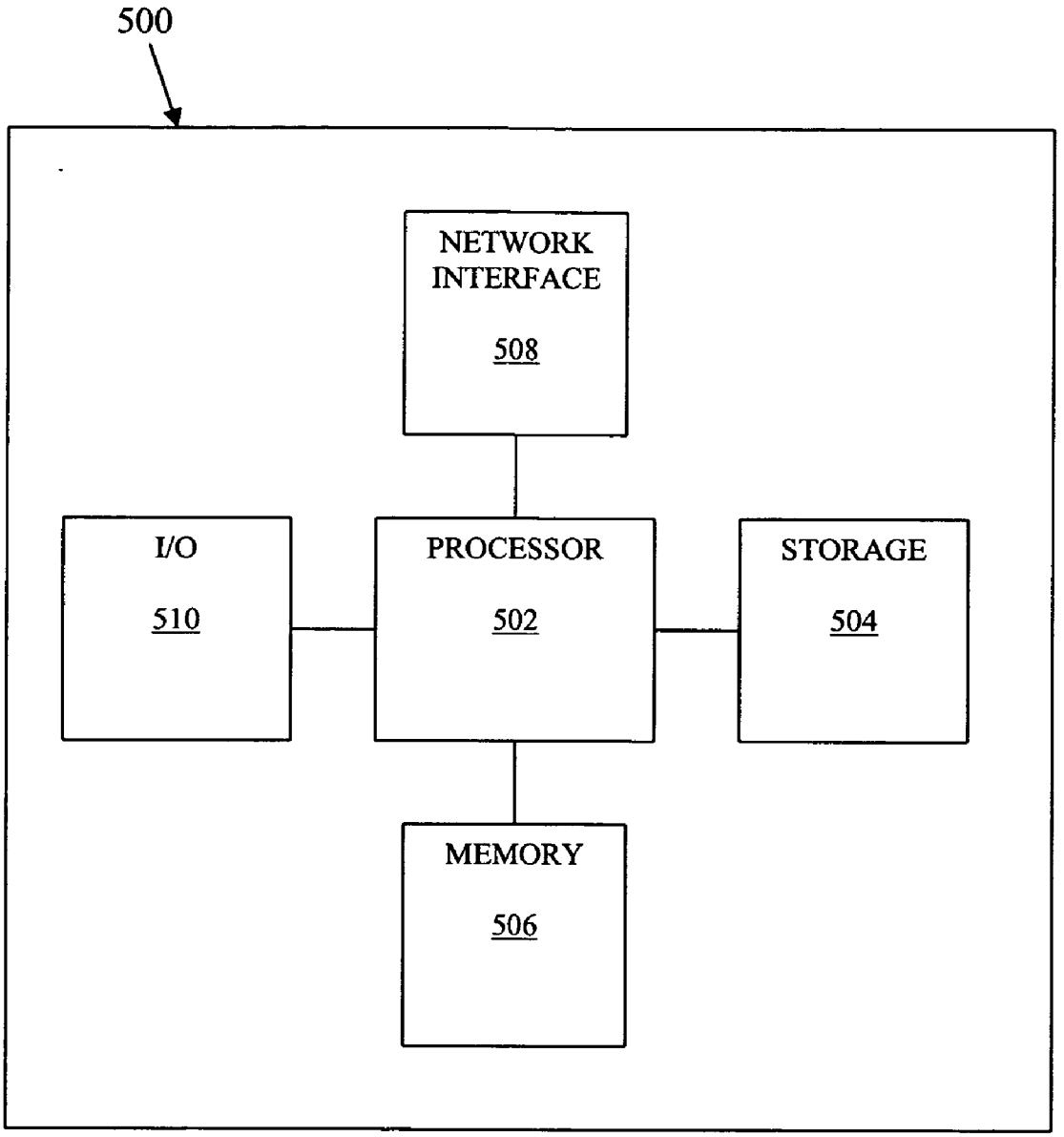

500
NETWORK INTERFACE
508
I/O
510
PROCESSOR
502
STORAGE
504
MEMORY
506

FIG. 5

DESIGN OF MULTI-USER DOWNLINK LINEAR MIMO PRECODING SYSTEMS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/870,930 filed on Dec. 20, 2006, which is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] The present invention is related generally to wireless transmission, and in particular, to throughput optimization of multi-user downlink linear multi-input multi-output (MIMO) precoding systems with quantized feedback.

[0003] In MIMO wireless transmission systems, higher rates of throughput (e.g., sum rates) for the overall system are desired. To facilitate gains in throughput, precoding may be used. Precoding is a generalized beamforming scheme for supporting multi-layer transmission in MIMO systems. In precoding, the multiple streams of the signals are emitted from the transmit antennas with independent and appropriate weighting so as to increase the link throughput at the receiver output.

[0004] In current systems, precoding algorithms for multi-user MIMO use either linear or nonlinear precoding algorithms. Linear precoding may achieve reasonable performance with lower complexity than nonlinear precoding. Nonlinear precoding may achieve greater capacity, but is very complex and accordingly slower. In general, nonlinear precoding is designed based on the concept of dirty paper coding (DPC). That is, known interference at the transmitter is subtracted without the penalty of resources if the optimal preceding scheme is applied to the transmission signal.

[0005] In general, multi user (MU-) MIMO systems offer significant gains over single user (SU-) MIMO schemes. In particular, by increasing the number of users there is a potential gain in system throughput. Even for a relatively small number of users, MU-MIMO systems can still outperform SU-MIMO systems by using an appropriate transmission scheme (such as precoding) for low- and mid-range SNRs.

[0006] With full channel state information at the transmitter (CSIT), the capacity of a Gaussian MU-MIMO channel is achieved by using DPC. However, there are insufficient solutions regarding capacity without CSIT or with partial CSIT. Many prior systems have instead used simpler transmission schemes with linear precoding. With partial CSIT, different schemes have been used. For example, a simple random beamforming algorithm has been used to realize the asymptotical gain of scheduling at the limit of large number of users. In another example, a robust opportunistic user scheduling and power allocation strategy based on limited feedback is used. However, these systems have sub-optimal throughput over the network because they do not have full CSIT and cannot quickly and efficiently schedule multiple users.

[0007] Therefore, there remains a need to improve throughput in MU-MIMO networks using CSIT.

BRIEF SUMMARY OF THE INVENTION

[0008] The present invention is generally directed to transmission in downlink multi-user MIMO networks. To maximize throughput over the network, quantized CSIT is sent through a low-rate feedback link feedback from a plurality of users back to a base station. The base station then determines a subset of the plurality of users to transmit one or more signals to based on the received feedback and determines a precoding matrix based on the received feedback from the plurality of users wherein the preceding matrix maximizes a sum-rate throughput for the subset of the plurality of users.

[0009] Additionally, based on the received feedback, the base station designs a quantization codebook. This codebook may be designed off-line and/or online. The codebook and/or precoding matrix are used to transmit signals to the users.

[0010] These and other advantages of the invention will be apparent to those of ordinary skill in the art by reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] FIG. **1** depicts an exemplary network in which the present invention may be implemented;

[0012] FIG. **2** depicts a flowchart of a method of transmission over a network;

[0013] FIG. **3** depicts a flowchart of a method of designing a quantization codebook according to an embodiment of the invention;

[0014] FIG. **4** is a flowchart of a method of designing a precoder according to an embodiment of the invention; and

[0015] FIG. **5** is a schematic drawing of a controller.

DETAILED DESCRIPTION

[0016] The present invention is related generally to downlink wireless networks, and in particular to throughput optimization in multi-user downlink linear multi-input multi-output (MIMO) precoding systems with quantized feedback. In general, multi-user MIMO (MU-MIMO) systems may offer significant gains over single user MIMO (SU-MIMO) schemes. In particular, by increasing the number of users (e.g., receivers) there is a potential gain in transmission scheduling. Even for a relatively small number of users, MU-MIMO systems may still outperform SU-MIMO systems by using an appropriate transmission scheme, such as preceding, for the low- and mid-range signal to noise ratios (SNRs).

[0017] To improve scheduling, linear precoding may be used in MU-MIMO systems where the quantized channel state information at the transmitter (CSIT) is provided through a low-rate feedback link. With full CSIT, the capacity of a Gaussian MU-MIMO channel may be achieved by using dirty-paper coding. Thus, in a linear precoder design—given quantized CSIT—the actual channels lie in a "neighborhood" of the quantized channel with certain distribution. Since the precoding is based on a specific quantization codebook, the performance of MU-MIMO schemes with quantized feedback using the inventive linear precoder design is directly affected by the quantization codebook. Similarly, a quantization codebook may be designed based on a capacity measure for general correlated channels. A new distance metric on the Grassmanian manifold is used to design the optimal precoder codebook. In application, this new metric is equivalent to the chordal distance at low SNRs and to the Fubini-Study distance at high SNRs. Thus, codebooks based on the inventive distance metric outperform codebooks designed based on chordal distance and Fubini-Study distance.

[0018] Further, transmissions to multiple users (e.g., receivers, reception antennas, etc.) may be scheduled to maximize throughput in wireless networks. The scheduling uses the quantized feedback from the users to create a precoder and/or a codebook for use in signal transmission.

[0019] FIG. **1** depicts an exemplary network **100**. Network **100** includes a base station **102** which has one or more (e.g., 1, 2, . . . , M) transmission antennas **104***a*, **104***b*, . . . , **104**M. The base station **102** may transmit signals via transmission antennas **104***a*-**104**M to one or more (e.g., 1, 2, . . . , K) receivers **106***a*, **106***b*, . . . , **106**K which may each have one or more reception antennas **108***a*, **108***b*, **108***c*, **108***d*, **108***e*, . . . , **108**N. Each of the receivers **106***a*-**106**K may be capable of transmitting feedback via respective feedback links **110***a*, **110***b*, . . . , **110**K.

[0020] Network **100** may be a downlink multi-user multiple input and multiple output (MU-MIMO) system. That is, network **100** may include multiple base stations **102** and/or multiple receivers **106***a*-**106**K and base stations **102** may transmit signals to the receivers **106***a*-**106**K.

[0021] Base station **102** may have and/or be a wireless transmitter (e.g., cellular site, satellite, Tx, etc.) as is known and accordingly may include a precoder **112** for precoding signals transmitted by transmission antennas **104***a*-**104**M and/or a scheduler **114** for scheduling one or more transmissions. In some embodiments, base station **102** may comprise multiple transmitters, each with multiple precoders **112** and transmission antennas **104***a*-**104**M. Further, base station **102** may comprise one or more reception antennas and/or inputs (not shown) for receiving input data and/or feedback signals, etc. and one or more controllers (e.g., controller **500** of FIG. **5**) and/or or appropriate control circuitry, symbol generators, multiplexers, etc.

[0022] Receivers (e.g., users) **106***a*-**106**K may be wireless reception devices (e.g. mobile telephone, ground station, Rx, etc.) as are known and may be capable of receiving and decoding (e.g., demultiplexing) signals received at reception antennas **108***a*-**108**N as well as transmitting feedback (e.g., quantized feedback) to base station **100**. It is understood that this feedback transmission may be wired and/or wireless over feedback links **110***a*-**110**K and/or may be transmitted using a combination of reception antennas **108***a*-**108**N (e.g., to transmit) and/or transmission antennas **104***a*-**104**M (e.g., to receive) and feedback links **110***a*-**110**K are shown in this manner for diagram simplicity. Feedback links **110***a*-**110**K may be low-rate feedback links providing channel state information to the transmitter. In this way, the transmitter may have (e.g., may determine from information received over feedback links **110***a*-**110**K) CSIT.

[0023] Precoder **112** and/or controller **500** may be capable of performing precoding functions. That is, the precoder **112** and/or controller may design, determine, and/or implement a precoder and/or a codebook for use in signal transmission by base station **102**.

[0024] In network **100**, the complex baseband signal model for the kth receiver (e.g., of receivers **106***a*-**106**K) is $y_k = H_k x + w_k$ where x is the M×1 transmitted signal vector, H_k is the $N_k \times M$ channel matrix, $w_k \sim N_c(0, I)$ is a circularly symmetric complex additive white Gaussian noise vector, and y_k is the $N_k \times 1$ received signal vector. In a block fading channel model in which the channel remains constant during the transmission of each packet (or codeword of length T) and the channel changes independently from one block to another, where the distribution of the channel state is known a priori, the average power constraint is given by $E[x^H x] \leqq P$.

[0025] For maximizing the sum-rate (e.g., the total rates from the base station **102** to active receivers **106***a*-**106**K) throughput in network **100**, the following assumptions are made. First, quantized channel state information for each receiver **106***a*-**106**K is provided via a feedback link **110***a*-**110**K. Second, only linear precoding is allowed by the base station **102**. Third, receivers **106***a*-**106**K do not decode other receivers' signals; they treat those signals as noise.

[0026] FIG. **2** depicts a flowchart of a method **200** of transmission over the network **100**. In at least one embodiment, transmission may be transmission of signals (e.g., wireless signals, information, data, etc.) from base station **102** to receivers **106***a*-**106**K. The method **200** begins at step **202**.

[0027] In step **204**, a channel is quantized and feedback is sent from receivers **106***a*-**106***k* to base station **102** over feedback links **110***a*-**110**K. In at least one embodiment, the channel being quantized may be indicative of a channel received from the base station **102** and information indicative of the channel may be quantized by receivers **106***a*-**106**K and/or base station **102**. The quantized channels may then be fed back as feedback to the base station **102** via feedback links **110***a*-**110**K.

[0028] For quantization of a channel, let $U_k D_k V_k^*$ be a singular value decomposition (SVD) of the kth receiver channel H_k. With B bits of feedback per receiver, the receivers **106***a*-**106**K quantize the first N columns of V_k (where $N \leqq \min(N_1, \ldots, N_k, M)$ is a fixed number predetermined by the base station **102**) using a quantization codebook $Q = \{Q_1, Q_2, \ldots, Q_2 B\}, Q_1 \in \mathcal{U}_k^{M \times N}$ as

$$\hat{V}_k(1{:}N) = \underset{Q \in Q}{\operatorname{argmin}} d(V_k(1{:}N), Q),$$

where d(.,.) is a distance metric. The columns of the quantized precoding matrix $\hat{V}_k(1{:}N)$ correspond to possible different streams for a particular receiver **106***a*-**106**K. Codebook design and choice of an appropriate distance metric are discussed in further detail below with respect to method **300**.

[0029] In step **206**, transmissions are scheduled. That is, at a given time, a determination may be made as to which receivers **106***a*-**106**K should receive transmissions. In some embodiments, based on the quantized feedback sent from the receivers **106***a*-**106**K to the base station **102** in step **204**, the base station **102** selects a subset of L, $L \leqq M$, streams (u) to transmit over transmission antennas **104***a*-**104**M.

[0030] In system **100** with K receivers **106***a*-**106**K where each receiver **106***a*-**106**K feeds back quantized channel state information (CSI) corresponding to the first N dominant eigenvectors $V_k(1{:}N)$ of its channel H_k, scheduling comprises choosing a set of L streams out of KN possible streams that have the best throughput performance. Thus, the optimum receiver subset selection requires calculating the performance for

$$\binom{KN}{L}$$

subsets and the computational complexity of finding the optimal precoder with maximum throughput for each subset is high. Hence, optimal scheduling is computationally intensive.

[0031] In some embodiments, an alternative scheduling criterion may be used. In such embodiments, the scheduler **114**

or another component (e.g., base station **102**, controller **500** of FIG. **5**, etc.) determines the subset of receivers **106***a*-**106**K that minimizes

$$\phi(\underline{\hat{V}}) = \sum_{i=1}^{L} \sum_{j=1, j\neq i}^{L} |\hat{v}_i^* \hat{v}_j|^2$$

for all matrices $\underline{\hat{V}}=[\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_L]^T$ composed of L out of KN quantized eigenvectors. To solve this scheduling problem, a numerical approach based on simulated annealing may be used where different subsets of quantized eigenvectors constitute the set of states and the energy of the state $\underline{\hat{V}}$ is given by $\phi(\underline{\hat{V}})$. In the ith iteration, the transition probability between the current state $\underline{\hat{V}}^{(i)}$ and the candidate state $\underline{\hat{V}}^{(i+1)}$ is given by

$$P\left(\underline{\hat{V}}^{(i)} \rightarrow \underline{\hat{V}}^{(i+1)}\right) = \min\left(1, \exp\left[\frac{1}{T_i}\left(\phi(\underline{\hat{V}}^{(i)}) - \phi(\underline{\hat{V}}^{(i+1)})\right)\right]\right)$$

where T_i is a sequence of decreasing positive real numbers.

[0032] In step **208**, a precoder (e.g., a precoding matrix) is found. The precoder may be found as described below with respect to method **400** and FIG. **4**.

[0033] At step **210**, the signals to be transmitted from base station **102** are precoded. Here, a quantization codebook may be designed (as described below with respect to FIG. **3** and method **300**) and/or a predetermined codebook may be used (e.g., selected, via a look-up, using controller **500**, etc.).

[0034] A transmitted signal x from the base station **102** may comprise L data streams, $u_1, u2, \ldots, u_L$, sent through column vectors $g_1, g_2, \ldots, g_L$ of a linear precoder G (e.g., precoder **112**). In this way,

$$x = Gu = \sum_{i=1}^{L} u_i g_i$$

where one or more streams may be intended for a receiver **106***a*-**106**K.

[0035] In some embodiments, steps **204**-**210** are performed jointly (e.g., in parallel) to find the optimal transmission scheme. In alternative embodiments, a decoupled (e.g., step by step) approach may be used where the quantization codebook design depends only on maximizing a single-receiver capacity, the scheduling is based on minimizing the sum of the inner products between the channel directions of all pairs of the scheduled receiver streams, and the optimal linear precoder is designed for the scheduled subset of the receiver streams based on maximizing the sum rate.

[0036] Following precoding in step **210**, the signals are transmitted by base station **102** to one or more of receivers **106***a*-**106**K in step **212**. The method ends at step **214**.

[0037] FIG. **3** depicts a flowchart of a method **300** of designing a quantization codebook according to an embodiment of the invention. A quantization codebook may be designed using a criterion based on a capacity measure. This differs from conventional codebook design (e.g., using the maximum mutual-minimum-distance). Thus, optimal codebook design may be based on the capacity measure, which is generally SNR dependant. While the capacity measurement does not directly define a valid distance metric for the design, a simple bound on the capacity expression can be used to formulate a quantization codebook design.

[0038] Since codebook design is independent of a receiver's index, for notational simplicity the subscript "k" will not always be used hereinafter. For example, the channel matrix of the kth receiver would be noted H. With B bits of feedback, designing a codebook may comprise finding the set $Q=\{Q_1, Q_2, \ldots, Q_{2^B}\}$ of semi-unitary matrices and mapping Q(H) from the set of channel conditions H to Q that maximizes the instantaneous information rate

$$C = \text{logdet}\left(I + \frac{P}{N} HQ(H)Q(H)*H*\right)$$

where Q(H) represents the M×N transmission matrix used for the channel condition H and Q(H)*Q(H)=I.

[0039] The method starts at step **302**. In step **304**, an evaluation set of channels (e.g., $\{H^{(l)}\}_{l=1}^{Size}$) is generated. In at least one embodiment, the evaluation set may be based on channel distributions. In step **306**, the evaluation set is factorized. That is, the evaluation set of channels from step **304** are singular value decomposed (e.g., $H^{(l)}=U^{(l)}D^{(l)}V^{(l)}$).

[0040] For further notational simplicity, Q(H) is represented only as Q hereinafter. Substituting the SVD of H=UDV* into

$$C = \text{logdet}\left(I + \frac{P}{N} HQ(H)Q(H)*H*\right),$$

$$C = \text{logdet}\left(I + \frac{P}{N} |D|^2 V^* QQV^*\right).$$

An approximate upper bound of the capacity may be expressed, where channel realizations have bounded norms (e.g., $\|H\|^2_F=\text{tr}[H^*H]=\text{tr}[D^2]\leqq\beta$), as

$$C \underset{\approx}{<} \text{logdet}\left(I + \frac{P}{N}\beta V^* QQV^*\right).$$

Considering the low SNR approximation (e.g., P→0),

$$C \underset{\approx}{<} \frac{P\beta}{N}\|V^* Q\|_F^2.$$

Similarly, considering the high SNR approximation (e.g., P→∞),

$$C \underset{\approx}{<} \log\left(\frac{P\beta}{N}|V^* Q|^2\right).$$

[0041] For 0<SNR<∞, since the transmitted signals are the linear combination of the columns of the precoder Q, the quantization codebook design relies on a metric that measures the distance between the subspaces spanned by different precoders (e.g., Q_1, and Q_2). Maximizing

$$C \underset{\approx}{<} \log\det\left(I + \frac{P}{N}\beta V^* Q Q V^*\right)$$

is related to minimizing the p-metric

$$d_p(Q, V) = \arccos\left[\frac{\left(I + \frac{P}{N}\beta V^* Q Q V^*\right)}{(1+p)^N}\right]$$

between the subspaces defined by V and Q on the Grassmanian manifold G(M,N) (e.g., the space of all N dimensional subspaces of an M dimensional vector space). The design of a codebook as discussed herein is based on the p-metric and accordingly depends on the SNR level. Therefore, based on the average power P, the value of the parameter p is chosen as

$$p = \frac{\beta P}{N}$$

where the resulting codebook is found where SNR varies in a neighborhood of the design SNR. The choice of β is discussed in further detail below. Of course, it follows that

$$p = \frac{SNR_{ave}}{N}\max_k\left\{\|H^{(l)}\|_F^2\right\} \text{ and } d_p(Q, V) = \arccos\left[\frac{\det(I + pV^* QQ * V)}{(1+p)^N}\right]$$

(e.g., the inventive distance metric).

[0042] In step **308**, the codebook $Q(0)=\{Q_1(0),Q_2(0), \ldots, Q_{2^B}(0)\}$ is initialized. In at least one embodiment, the codebook is initialized randomly. In step **310**, partitions $\{\mathcal{V}_k, k=1, 2, \ldots, 2^B\}$ are found where $\mathcal{V}_k$ is the set of all right singular vectors defined in the partitioning discussed below.

[0043] Given the quantization codebook $Q=\{Q_1,Q_2, \ldots, Q_2^B\}$, the optimal partitioning of the quantizer Q(H) satisfies:

$$\mathcal{V}_k=\{V\in\mathbb{C}^{M\times N}: d_p(V,Q_k)\leq d_p(V,Q_j), \forall j\neq k.$$

[0044] In step **312**, the centroid is found. In at least one embodiment, the centroid may be expressed as $\{Q_k(i), k=1, 2, \ldots, 2^B\}$. To satisfy the centroid condition,

$$Q_k = \arg\min_{Q\in\mathbb{C}^{M\times N}, Q^*Q=1} E[d_p(Q, V) \mid V \in V_k]$$

may be solved for the kth partition. In at least one embodiment, this may be solved numerically. To employ a gradient-descent search algorithm, real representations of the matrices Q,H,V, are used as $\overline{Q},\overline{H},\overline{V}$, respectively, defined by

$$\overline{Q} = \begin{bmatrix} \Re e(Q) & -\Im m(Q) \\ \Im m(Q) & \Re e(Q) \end{bmatrix}.$$

[0045] The semi-unitary matrix $Q\in\mho_k^{M\times N}$ is then parameterized using $N_\Phi=2(M-N+1)N$ independent real parameters ϕ_i, $I=1, \ldots, N_\Phi$. Therefore,

$$Q_k = \arg\min_{Q\in\mathbb{C}^{M\times N}, Q^*Q=1} E[d_p(Q, V) \mid V \in V_k]$$

may be determined by considering the unconditional problem in terms of the vector $\phi=[\phi_1, \phi_2, \ldots, \phi_{N\Phi}]$ This leads to the objective function

$$\min_\Phi J(Q(\Phi)) = \min_\Phi E[d_p(\overline{Q}(\Phi), \overline{V} \mid V \in \mho_k].$$

The derivative of this objective function with respect to ϕ_k is then

$$\frac{\partial}{\partial\phi_k} J(\overline{Q}(\Phi)) = tr\left\{\left[\frac{\partial\overline{Q}(\Phi)}{\partial\phi_k}\right]^T \nabla_{\overline{Q}} J(\overline{Q})\right\}, 1 \leq k \leq N_\Phi$$

where

$$\nabla_{\overline{Q}} J(\overline{Q}) = E\left[-\frac{(1-F^2)^{-1/2}}{(1+p)^N}\nabla_{\overline{Q}} \det\left(I + p\overline{V}^T\overline{Q}\overline{Q}^T\overline{V}\right) \middle| V \in \mho_k\right] =$$

$$\nabla_{\overline{Q}} J(\overline{Q}) = E\left[-2\,\mathrm{pF}(1-F^2)^{-1/2}\overline{V}\overline{V}^T\overline{Q}\left(I + p\overline{V}^T\overline{Q}\overline{Q}^T\overline{V}\right)^{-1} \mid V \in V_k\right]$$

with

$$F = \left[\frac{\det\left(I + p\overline{V}^T\overline{Q}\overline{Q}^T\overline{V}\right)}{(1+p)^N}\right].$$

Therefore, a gradient descent search algorithm may be used with the recursion $\Phi^{(l+1)}=\Phi^{(l)}-\mu_l\nabla_\Phi J(\overline{Q}(\Phi))|_{\Phi^{(l)}}$ where μ_l is the sequence of step sizes.

[0046] In step **314**, a check is performed to determine if the centroid $\Phi^{(l+1)}=\Phi^{(l)}-\mu_l\nabla_\Phi J(\overline{Q}(\Phi))|_{\Phi^{(l)}}$ has converged. Here, (is used to find the unitary matrix $Q(\phi)$ and $Q(\phi)$ is the centroid for $\mathcal{V}_k$ described above. If the centroid has not converged, step **312** is repeated iteratively until convergence.

[0047] If the centroid has converged, the method proceeds to step **316** and a check is performed to determine if the quantization codebook $\mathcal{Q}=\{Q_k\}$ has converged. If the codebook has converged, the method proceeds to step **318** and the quantization codebook $\mathcal{Q}$ is determined. That is, when the quantization codebook $\mathcal{Q}$ converges after a number of iterations, the set of all Q_k is the codebook. If the codebook has not converged, the method returns to step **310**.

[0048] In actual implementation, this algorithm converges readily. Accordingly, the method **300** may be used in on-line (e.g., live, continuous transmission, etc.) applications. The method ends at step **318**.

[0049] FIG. **4** is a flowchart of a method **400** of designing a precoding matrix and for precoding and transmission according to an embodiment of the invention. In some embodiments, the precoding matrix (alternatively referred to as the "precoder") may be stored on, designed by and/or implemented by precoder **112** of FIG. **1** and may be used, for precoding transmissions in network **100**. In the same or alternative

embodiments, the precoding matrix may be designed for transmission to multiple receivers **104***a*-**104**K based on quantized feedback in order to maximize the sum-rate. The method **400** begins at step **402**.

[0050] In step **404**, each receiver **106***a*-**106**K performs SVD on its incoming channel (e.g., $H_k=U_kD_kV_k^*$) and quantizes the dominant right eigenvectors $V_k(1{:}N)$. In step **406**, the resultant indices are fed back to the base station **102** via feedback links **110***a*-**110**K. The scheduler **114** or another component (e.g., controller **500** of FIG. **5**, etc.) may select L transmission streams as discussed above with respect to step **206** of method **200**. $\underline{\hat{V}}=[\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_L]^T$ denotes the matrix of the quantized eigenvectors of the L scheduled receiver streams. In general, $\hat{v}_k$ corresponds to the kth receiver **106***a*-**106**K for which the channel matrix is given by H_k. In some embodiments, streams belonging to the same receiver are jointly detected. In this way, if $\hat{v}_i$ and $\hat{v}_j$ belong to the same receiver **106***a*-**106**K, $H_i=H_j$. At the transmitter (e.g., base station **102**), each set of quantized indices restricts the set of channel conditions for the scheduled receivers to a Voronoi region: $D(\hat{V})=\{(H^1, \ldots H_L):\|v_k-\hat{v}_k\|\leqq\|v_k-q_j\|, \hat{v}_k,q_j\epsilon Q, 1\leqq j\leqq 2^B, \hat{v}_k\neq q_j$, and k=1, 2, . . . , L} and $d(v_k, \hat{v}_k)\leqq d(v_k,q_j)$ where d(.,.) is the distance (e.g., design) metric described above, $v_1, v_2, \ldots, v_L$ are the corresponding right eigenvectors of the channel matrices $H_1, H_2, \ldots, H_L$ to the quantized eigenvectors $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_L$, respectively, and $Q=\{q_1,q_2,\ldots,q_2^B\}$ is the quantization codebook as discussed above.

[0051] In step **406**, the data streams are precoded. In some embodiments, L data streams $u_1, u_2, \ldots, u_L$ may be precoded by the M×L precoding matrix $G=[g_1, g_2, \ldots, g_L]$ with $g_i\epsilon C^{M\times 1}$. The received signal by the kth receiver **106***a*-**106**K is

$$y_k = H_k g_k u_k + \sum_{i=1,i\neq k}^{L} H_k g_i u_i + w_k.$$

By treating interference between receivers **106***a*-**106**K as noise, the rate of the kth stream is

$$R_k(\underline{\hat{V}}) = \log\left(1+(H_k g_k)*\left[1+\sum_{i=1,i\neq k}^{L}(H_k g_i)(H_k g_i)^*\right]^{-1}(H_k g_k)\right).$$

[0052] Channel knowledge at the base station **102** for each set of feedback indices defined by the set $H\epsilon D(\underline{\hat{V}})$ is a deterministic function of the channel state information at the receiver (CSIR), the sum rate throughput may be

$$R_{sum} = \max_{g}\sum_{\hat{V}} Pr\{H\in D(\underline{\hat{V}})\}E_{H\in D(\underline{\hat{V}})}\left[\sum_{k=1}^{L} R_k(\underline{\hat{V}})\right]$$

where optimization is taken over the set of precoding matrices $G=\{G(\underline{\hat{V}}),\forall\underline{V}\}$ for each Voronoi region.

[0053] The average power used by each precoder $G\epsilon G$ may then be $E[|Gu|^2]=tr(GG^*)$ when $E[uu^*]=I$. Therefore, G is defined such that the total power constraint $\Sigma_{\hat{V}}Pr(H\epsilon D(\underline{\hat{V}}))tr(G(\underline{\hat{V}})G(\underline{\hat{V}})^*)\leqq P$ is satisfied. In alternative embodiments, constant power may be used. In such cases, the power constraint may be $tr(G(\underline{\hat{V}})G(\underline{\hat{V}})^*)\leqq P$.

[0054] In step **408**, the rate may be optimized. That is, the optimization problem R_{sum} may be solved. In some embodiments, the optimization may be solved separately for each Voronoi region. For a generic Voronoi region D and a precoder G, the optimization problem may be:

$$\max_G J = E_{H\in D}\left[\sum_{k=1}^{L}\log\left(1+(H_k g_k)*\left[I+\sum_{i=1,i\neq k}^{L}(H_k g_i)(H_k g_i)^*\right]^{-1}(H_k g_k)\right)\right]$$

such that $tr(GG^*)\leqq P$.

[0055] In some embodiments, a gradient descent search algorithm may be used to find the solution. The gradient of the objective function with respect to the precoder G may be determined analytically. The $(l,j)^{th}$ element of the objective function in such instances is

$$\frac{\partial J}{\partial g_{lj}} = E_{H\in D}\left[\frac{2}{A_l}(H_l g_l)^*\sum_{l}^{-1} h_j^{(1)} + \sum_{k=1,k\neq l}^{L}\frac{2}{A_k}(H_k g_k)^*\left[-\sum_{k}^{-1} h_j^{(k)}(H_k g_l)^*\sum_{k}^{-1}\right](H_k g_k)\right]$$

where

$$\sum_{k} = I+\sum_{i=1,i\neq k}^{L}(H_k g_i)(H_k g_i)^*,\ A_k = 1+(H_k g_k)^*\sum_{k}^{-1}(H_k g_k),\ h_j^{(k)}$$

is the jth row if H_k and g_{lj} is the jth element of the column vector g_l. The objective function may then be expressed as $\nabla_G J=E_{H\epsilon D}[H^*\psi^*A+H^*\psi^*\Upsilon A-H^*\psi^*A\psi HG]$ where

$$A = diag\left(\frac{1}{A_1}, \frac{1}{A_2}, \ldots, \frac{1}{A_L}\right),$$

$H=[H_1{}^TH_2{}^T \ldots H_L{}^T]^T$, $\psi=diag(T_1, T_2, \ldots, T_L)$, $\Upsilon=diag(T_1H_1g_1, T_2H_2g_2, \ldots, T_2H_Lg_L)$, and

$$T_l = (H_l g_l)^*\sum_{l}^{-1},\ 1\leq l\leq L.$$

[0056] In step **410**, the precoding matrix may be determined. In some embodiments, the precoder (precoding matrix) may be determined (e.g., the optimal G may be found) using a gradient descent algorithm as is known with the following recursion:

$$\tilde{G}^{k+1} = G^{(k)} - \mu_k\nabla_G J(G^{(k)}),$$

-continued

$$G^{(k+1)} = \frac{\sqrt{P}\,\tilde{G}^{(k+1)}}{\sqrt{tr(\tilde{G}^{(k+1)}\tilde{G}^{(k+1)*})}}$$

[0057] where $G^{(k)}$ is the precoding matrix after the kth iteration and u_k is the step size parameter. Normalization with

$$G^{(k+1)} = \frac{\sqrt{P}\,\tilde{G}^{(k+1)}}{\sqrt{tr(\tilde{G}^{(k+1)}\tilde{G}^{(k+1)*})}}$$

may be performed to satisfy the power constraint described above.

[0058] Step **410** may be initialized to an arbitrary semi-unitary matrix G that satisfies the power constraint tr(GG*) ≦P. In the same or alternative embodiments, step **410** may be initially started at the solution to

$$\tilde{G}^{(0)} = \underline{V}^{*}\left[\underline{\hat{V}}\,\underline{V}^{*} + \frac{L}{P}I\right]^{-1}$$

which may facilitate faster convergence of the recursion.

[0059] In some embodiments, an alternative algorithm to the gradient descent algorithm may be used. For example, since the optimal solution satisfies $G=\mathbb{E}_{He\mathcal{D}}[(H^*\psi^*A\psi H)^{-1}H^*\psi^*(I+\Upsilon)A]$, an alternative iterative algorithm may be used. In such an algorithm, at the kth step, G may be calculated using $G_{(k)}$ (e.g., the precoding matrix at the kth iteration) and $G^{(k+1)}\,\mathbb{E}_{He\mathcal{D}}[(H^*\psi A\psi H)^{-1}H^*\psi^*(I+\Upsilon)A]|_{G^{(k)}}$. In this way, the new precoding matrix $G^{(k+1)}$ may be obtained after normalization. Other algorithms for determining the precoding matrix G may be used as appropriate.

[0060] The method ends at step **412**.

[0061] In order to detect their streams, receivers **106***a*-**106**K have to know the precoder used by the transmitter (e.g., base station **102**). This information is provided to the set of scheduled receivers **106***a*-**106**K through a reconfirmation process used by the base station **102**. This reconfirmation is done either by sending a dedicated precoded pilot for each user or by re-transmitting the index of the Voronoi region (e.g., the set of all scheduled user feedback indices). The size of precoding codebook is 2^{BL} for L streams and B bits of quantization of each eigenvector. Hence, although the precoding codebook can be obtained off-line based on a given quantization codebook, it might be practically not possible to store the precoding codebook because the size of the codebook could be very large. However, by implementing the same preceding method and the corresponding initialization as described above, the precoder may be computed on-line both at the base station **102** and at the receivers **106***a*-**106**K based on the quantization indices of the scheduled users.

[0062] The use of feedback may become undesirable at high SNR if a codebook with fixed size is employed since at high SNR a simple TDMA strategy without feedback will outperform a multi-user transmission strategy with imperfect CSIT. If the codebook size is fixed, transmission rank control may be employed by reducing the number of scheduled user streams as SNR increases. The sum-rate performance of such an adaptive transmission strategy is not limited at high SNR and is may improve performance over a TDMA strategy without CSIT and a multi-user precoding scheme with fixed transmission rank. The optimal transmission rank depends on the average SNR and can be computed off-line by using a simple greedy method.

[0063] Given the SNR, the precoder rank that results in the highest average sum-rate performance may be found. A precoder may be determined off-line as described above with respect to FIG. **4** for fixed transmission rank over different channel realizations. The rank that maximizes the average throughput performance may then be chosen. Therefore, in practical implementation a lookup table may be used with the available power and average SNR to determine the appropriate precoder rank that corresponds to the number of user streams to be selected by the scheduler **114**.

[0064] FIG. **5** is a schematic drawing of a controller **500** according to an embodiment of the invention. Controller **500** may be used in conjunction with and/or may perform the functions of base station **102**, precoder **112**, and/or scheduler **114**. In the same or alternative embodiments, controller **500** may reside at, be component of, and/or may be used by one or more receivers **106***a*-**106**K. That is, controller **500** may be adapted to perform functions of a transmitter and/or a receiver as is known.

[0065] Controller **500** contains a processor **502** which controls the overall operation of the controller **500** by executing computer program instructions which define such operation. The computer program instructions may be stored in a storage device **504** (e.g., magnetic disk, database, etc.) and loaded into memory **506** when execution of the computer program instructions is desired. Thus, applications for performing the herein-described method steps, such as preceding, codebook construction, transmitting and/or receiving data, and throughput optimization, in methods **200**, **300**, and **400** are defined by the computer program instructions stored in the memory **506** and/or storage **504** and controlled by the processor **502** executing the computer program instructions. The controller **500** may also include one or more network interfaces **508** for communicating with other devices via a network (e.g., a peer to peer network, etc.). The controller **500** also includes input/output devices **510** (e.g., display, keyboard, mouse, speakers, buttons, etc.) that enable user interaction with the controller **500**. Controller **500** and/or processor **502** may include one or more central processing units, read only memory (ROM) devices and/or random access memory (RAM) devices. One skilled in the art will recognize that an implementation of an actual controller could contain other components as well, and that the controller of FIG. **5** is a high level representation of some of the components of such a controller for illustrative purposes.

[0066] According to some embodiments of the present invention, instructions of a program (e.g., controller software) may be read into memory **506**, such as from a ROM device to a RAM device or from a LAN adapter to a RAM device. Execution of sequences of the instructions in the program may cause the controller **500** to perform one or more of the method steps described herein, such as those described above with respect to methods **200**, **300**, and **400**. In alternative embodiments, hard-wired circuitry or integrated circuits may be used in place of, or in combination with, software instructions for implementation of the processes of the present invention. Thus, embodiments of the present invention are not limited to any specific combination of hardware, firmware, and/or software. The memory **506** may store the

software for the controller **500**, which may be adapted to execute the software program and thereby operate in accordance with the present invention and particularly in accordance with the methods described in detail above. However, it would be understood by one of ordinary skill in the art that the invention as described herein could be implemented in many different ways using a wide range of programming techniques as well as general purpose hardware sub-systems or dedicated controllers.

[0067] Such programs may be stored in a compressed, uncompiled and/or encrypted format. The programs furthermore may include program elements that may be generally useful, such as an operating system, a database management system, and device drivers for allowing the controller to interface with computer peripheral devices, and other equipment/components. Appropriate general purpose program elements are known to those skilled in the art, and need not be described in detail herein.

[0068] The foregoing Detailed Description is to be understood as being in every respect illustrative and exemplary, but not restrictive, and the scope of the invention disclosed herein is not to be determined from the Detailed Description, but rather from the claims as interpreted according to the full breadth permitted by the patent laws. It is to be understood that the embodiments shown and described herein are only illustrative of the principles of the present invention and that various modifications may be implemented by those skilled in the art without departing from the scope and spirit of the invention. Those skilled in the art could implement various other feature combinations without departing from the scope and spirit of the invention.

1. A method of operation of a downlink multi-user MIMO system comprising:

receiving quantized feedback from a plurality of users;

determining a subset of the plurality of users to transmit one or more signals to based on the received feedback; and

determining a precoding matrix based on the received feedback from the plurality of users wherein the precoding matrix maximizes a sum-rate throughput for the subset of the plurality of users.

2. The method of claim **1** further comprising:

scheduling transmission to the plurality of users using the preceding matrix.

3. The method of claim **2** further comprising:

transmitting one or more signals over a plurality of channels to the plurality of users according to the transmission schedule.

4. The method of claim **1** further comprising:

determining a transmission signal based at least in part on the determined precoding matrix.

5. The method of claim **4** wherein determining the transmission signal comprises determining a result of x=Gu wherein:

x is the transmission signal;

G is the determined precoding matrix; and

u is a data stream to be transmitted.

6. The method of claim **1** wherein determining the precoding matrix comprises:

determining the precoding matrix with a gradient-descent search algorithm.

7. The method of claim **1** wherein determining the precoding matrix comprises:

determining the precoding matrix with an iterative algorithm comprising:

(a) initializing a precoding matrix $G^{(0)}$;

(b) determining a precoding matrix $G^{(k)}$ at a kth iteration;

(c) determining $G^{(k+1)}\mathbb{E}_{H\in D}[(H^*\psi^*A\psi H)^{-1}H^*\psi(I+\Upsilon)A]|_{G^{(k)}}$ wherein H is a channel matrix of transmission channels to the plurality of users, E is an average power of the channel matrix, D is a generic Voronoi region,

$$A = diag\left(\frac{1}{A_1}, \frac{1}{A_2}, \ldots, \frac{1}{A_L}\right),$$

$H=[H_1{}^T H_2{}^T \ldots H_L{}^T]^T$, $\psi=\mathrm{diag}(T_1,T_2,\ldots,T_L)$, $\Upsilon=\mathrm{diag}(T_1H_1g_1,T_2H_2g_2,\ldots,T_2H_Lg_L)$, and

$$T_l = (H_l g_l)^* \sum_{l}^{-1}, \; 1 \le l \le L; \text{ and}$$

iteratively repeating steps (b) and (c) until G converges.

8. The method of claim **1** wherein the quantized feedback is obtained by performing singular value decomposition on channels H_k corresponding to the plurality of users wherein $H_k=U_kD_kV_k^*$ and U is matrix of receive weights, D is a diagonal matrix of eigenvectors of H, V* is a matrix of transmit weights, and k is a number of users.

9. The method of claim **7** wherein the precoding matrix is determined off-line and further comprising:

determining from a look-up table a precoder rank at a base station based at least in part on an average channel condition and a transmit signal-to-noise ratio that results in the highest average sum-rate performance.

10. The method of claim **2** wherein scheduling transmission comprises:

determining a subset K of the plurality of users that minimizes

$$\phi(\underline{\hat{V}}) = \sum_{i=1}^{L} \sum_{j=1, j\neq i}^{L} |\hat{v}_i^* \hat{v}_j|^2$$

for all matrices $\underline{\hat{V}}=[\hat{v}_1,\hat{v}_2,\ldots,\hat{v}_L]^T$ composed of L user streams out of KN quantized eigenvectors.

11. The method of claim **1** wherein determining the precoding matrix comprises:

determining the preceding matrix with a gradient descent search algorithm comprising:

determining an $(l, j)^{th}$ element of an objective function given by

$$\frac{\partial J}{\partial g_{ij}} = E_{H\in D}\left[\frac{2}{A_l}(H_l g_l)^* \sum_{l}^{-1} h_j^{(l)} + \sum_{k=1,k\neq l}^{L} \frac{2}{A_k}(H_k g_k)^* \left[-\sum_{k}^{-1} h_j^{(k)}(H_k g_l)^* \sum_{k}^{-1}\right](H_k g_k)\right]$$

where

$$\sum_k = I + \sum_{i=1, i\neq k}^{L} (H_k g_i)(H_k g_i)^*,$$

$$A_k = 1 + (H_k g_k)^* \sum_k^{-1} (H_k g_k),\ h_j^{(k)} \text{ is a } j^{th}$$

row if H_k and g_{ij} is the jth element of the column vector g_i;

determining $\nabla_G J = E_{H\epsilon D}[H^*\psi^*A + H^*\psi^*\Upsilon A - H^*\psi^*a\psi HG]$ where

$$A = diag\left(\frac{1}{A_1}, \frac{1}{A_2}, \ldots, \frac{1}{A_L}\right),$$

$H=[H^T H_2^T \ldots H_L^T]^T$, $\psi=\text{diag}(T_1, T_2, \ldots T_L)$, $\Upsilon=\text{diag}(T_1H_1g_1, T_2H_2g_2^* \ldots, T_2H_Lg_L)$, and

$$T_l = (H_l g_l)^* \sum_l^{-1},\ 1 \le l \le L;$$

and

determining the precoding matrix with a gradient descent search algorithm using a recursion $G^{k+1}=G^{(k)}-\mu_k\nabla_G J(G^{(k)})$ where $G^{(k)}$ is the precoding matrix after the kth iteration and u_k is a step size parameter.

12. The method of claim **11** further comprising:

$$G^{(k+1)} = \frac{\sqrt{P}\,\tilde{G}^{(k+1)}}{\sqrt{tr(\tilde{G}^{(k+1)}\tilde{G}^{(k+1)*})}}$$

normalizing $G^{(k+1)}$ with to satisfy a power constraint.

13. The method of claim **12** wherein the power constraint is $\Sigma_{\underline{V}}\Pr(H\epsilon D(\underline{\hat{V}}))\text{tr}(G(\underline{\hat{V}})G(\underline{\hat{V}})^*)\leqq P$.

14. A method of designing a quantization codebook for a multi-user MIMO system comprising:

(a) generating an evaluation set of channels based on distributions of channels in the multi-user MIMO system;

(b) performing singular value decomposition on the evaluation set of channels;

(c) initializing the quantization codebook;

(d) determining partitions;

(e) determining a centroid; and

(f) iteratively repeating steps (d) and (e) until a predetermined number of iterations is reached.

15. The method of claim **14** wherein the partitions are $\{V_k, k=1, 2 \ldots, 2^B\}$.

16. The method of claim **14** wherein the centroid is $\{Q_k(i), k=1, 2, \ldots, 2^B\}$ and determining the centroid comprises:

determining $\mathfrak{V}_k=\{V\epsilon C^{M\times N}: d_p(V,Q_k)\leqq d_p(V,Q_j), \forall j\neq k$ wherein V is a transmission weight, $\mathbb{C}$ is the set of all complex numbers, N is the number of columns of V, and M is a number of reception antennas.

17. The method of claim **16** wherein

$$d_p(Q, V) = \arccos\left[\frac{\det(I + pV^*QQ*V)}{(1+p)^N}\right].$$

18. A method of data transmission in a downlink MIMO system comprising:

receiving quantized feedback from one or more users; and

selecting one or more precodes from a precoding matrix generated based on the received feedback wherein the preceding matrix maximizes a sum-rate throughput transmission rate for a subset of the one or more users.

19. The method of claim **18** further comprising:

determining a subset K of the plurality of users that minimizes

$$\phi(\underline{\hat{V}}) = \sum_{i=1}^{L} \sum_{j=1, j\neq i}^{L} |\hat{v}_i^* \hat{v}_j|^2$$

for all matrices $\underline{\hat{V}}=[\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_L]^T$ composed of L user streams out of KN quantized eigenvectors.

20. The method of claim **18** further comprising:

determining a transmission schedule for the subset K of the one or more users; and

transmitting one or more signals to the subset K of the one or more users according to the transmission schedule.

* * * * *